Statistical methods and data-driven models to predict glass properties

The case of glass melt viscosity

D. Perret, N. Bisbrouck, M. Neyret, C. Chabal

CEA Marcoule

DES/ISEC/DPME

Approaches for property modeling

- Theoretical, cognitive approach

  Based on our intrinsic knowledge of the phenomenon, on the fundamental laws of physics and chemistry (conservation of energy, momentum, equations of diffusion, thermodynamics,…)

- Empirical approach

  Based on a set of experimental data (*data-driven* models). Mathematical, statistical approach, which ignores any physicochemical knowledge of the phenomenon

- Mixed approach

  Combination of the two previous approaches

For these three classes of models, there are different types: linear or non-linear, static or dynamic, deterministic or stochastic, continuous or discrete,…

Approaches for property modeling

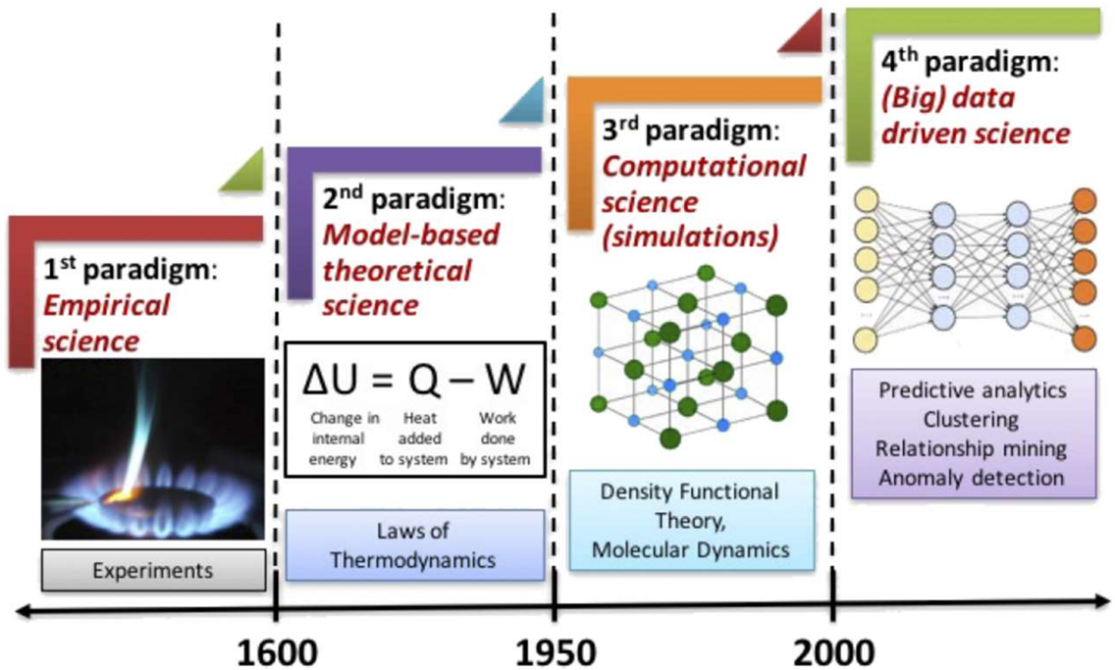


FIG. 1. The four paradigms of science: empirical, theoretical, computational, and data-driven.

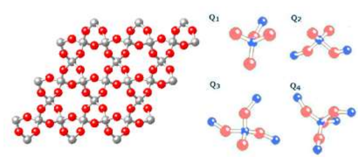
From A. Agrawala and A. Choudhary. APL Mater. 4, 053208 (2016); https://doi.org/10.1063/1.4946894

Approaches for property modeling

Property

Composition

Structure



- Degree of polymerization
- Topological constraints
- …

Forward models
Property prediction

Inverse models
Materials discovery

Approaches for property modeling

- First attempt for the calculation of glass properties from their composition proposed by Winckelmann and Schott at the end of the 19th century

- Theoretical Principle of Additivity

M.B. Volf*, Mathematical Approach to Glass,* Elsevier Science Publishers, 1988

$$G = \sum g(G)_i \, x_i$$

G is the property of the glass
$g(G)_i$ is the additive factor for oxide i and property G
x_i is the amount of oxide i

  - Generally valid when investigating suitably narrow composition range
  - Errors in additive calculation could be due to phase separation, crystallization, degree of cross-linking, anomalies in the cross-linked structure, interaction between ions,…

Statistical modeling of glass properties

« Design Of Experiments » (DOE) methodology

« Machine Learning » (ML) methodology

Statistical modeling of glass properties

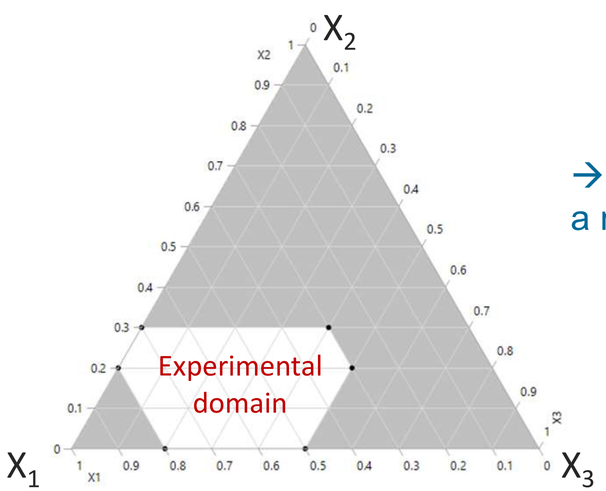
« Design Of Experiments » (DOE) methodology

« Machine Learning » (ML) methodology

Design of experiments methodology

General principle

"a branch of applied statistics that deals with planning, conducting, analyzing, and interpreting controlled tests to evaluate the factors that control the value of a parameter or group of parameters" *(from the American Society for Quality)*

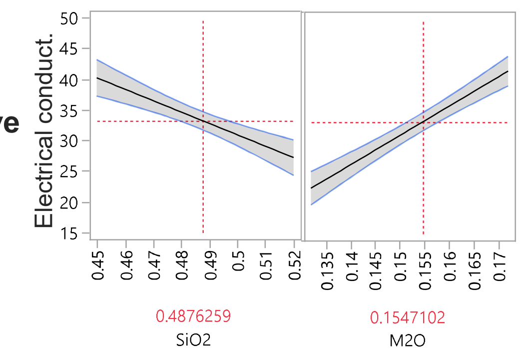


Mixture Design

"a statistical method to study cause-effect and phenomena-response relationships in processes and phenomena" *(Lazić, 2004)*

→ Acquire maximum knowledge from a minimum number of runs

Identification of **oxides that have a significant effect** on the response variation



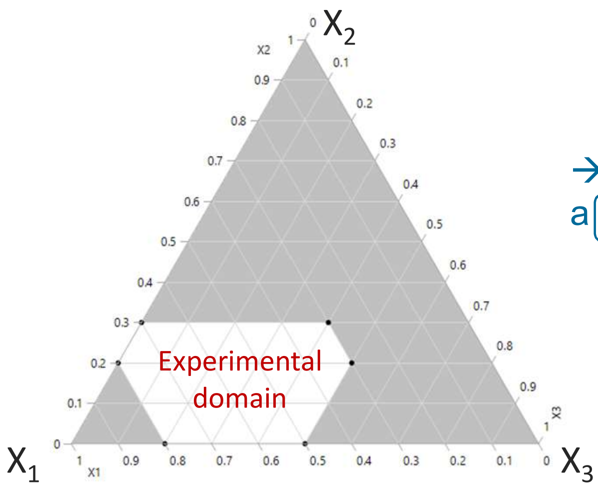
Predictive model for any location within the experimental domain

$$E(Y) = \sum_{i=1}^{p} a_i x_i + selected\left\{\sum_{i=1}^{p-1}\sum_{j=1}^{p} a_{ij} x_i x_j\right\}$$

Design of experiments methodology

General principle

"a branch of applied statistics that deals with planning, conducting, analyzing, and interpreting controlled tests to evaluate the factors that control the value of a parameter or group of parameters" *(from the American Society for Quality)*

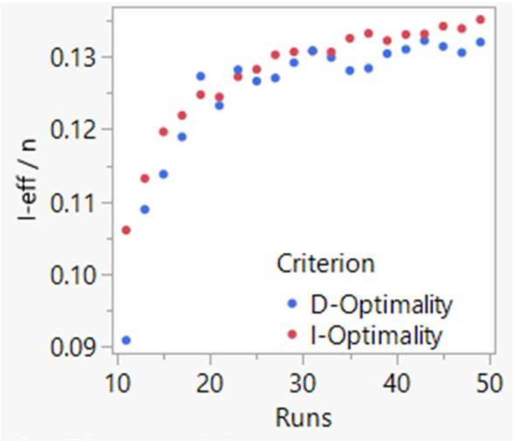


Mixture Design

"a statistical method to study cause-effect and phenomena-response relationships in processes and phenomena" *(Lazić, 2004)*

→ Acquire maximum knowledge from a minimum number of runs

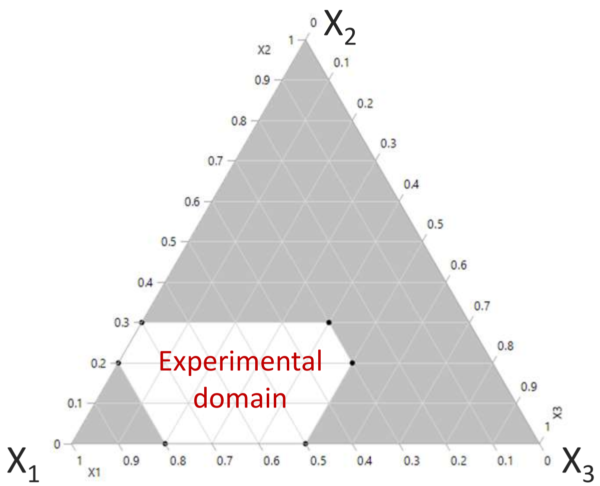
Evolution of **optimality criteria** as a function of the number of runs



→ about **20-25 experiments** for a mixture DOE with 8-10 oxides + relational constraints (Fleury 2014, Piovesan 2017)

Design of experiments methodology

General principle

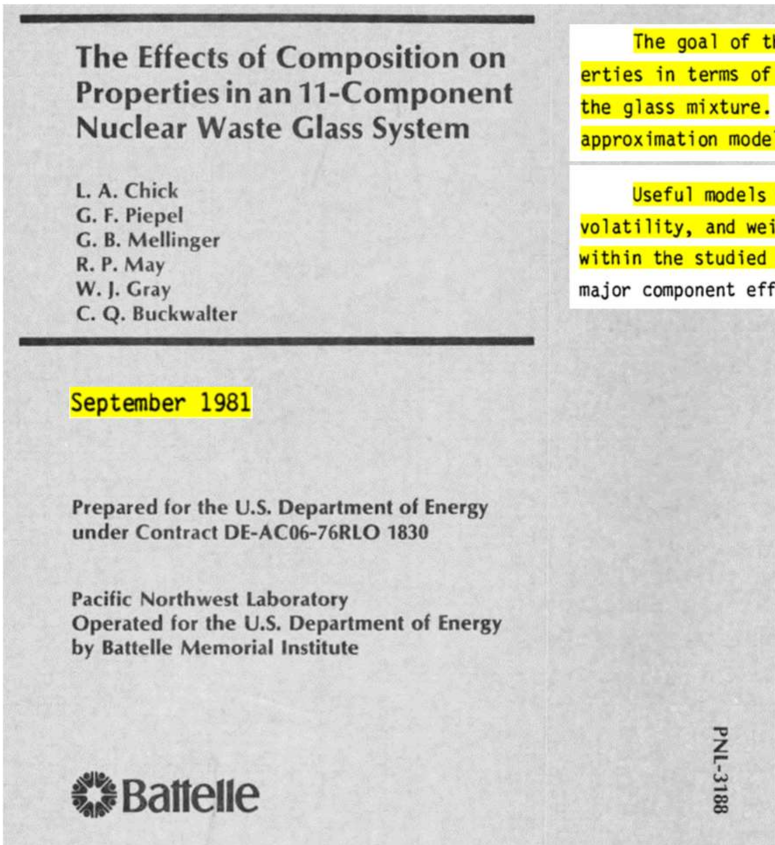


Mixture Design

- Key points:
  - Adequate composition domain boundaries (no possible extrapolation)
  - Definition of the most optimal number of runs
  - Best model selection (beware risk of overfitting)
  - Model validation (additional runs)
- Methodology can be applied to glasses containing up to 10-12 oxides
- Robust models, but no extrapolation outside the composition domain

PNNL legacy

From design of experiments to database models (1/2)



The Effects of Composition on Properties in an 11-Component Nuclear Waste Glass System

L. A. Chick
G. F. Piepel
G. B. Mellinger
R. P. May
W. J. Gray
C. Q. Buckwalter

September 1981

Prepared for the U.S. Department of Energy under Contract DE-AC06-76RLO 1830

Pacific Northwest Laboratory
Operated for the U.S. Department of Energy
by Battelle Memorial Institute

Battelle

PNL-3188

The goal of this study was to predict and understand various glass properties in terms of the mole percent (mol%) proportions of the components in the glass mixture. Due to the lack of theoretical models for complex systems, approximation models (see Section 2.2) were used. As the number of components

Useful models were developed for predicting crystallinity, viscosity, volatility, and weight loss-type chemical durability of compositions falling within the studied region. Figure 1 presents a qualitative summary of the major component effects on the successfully modeled properties.

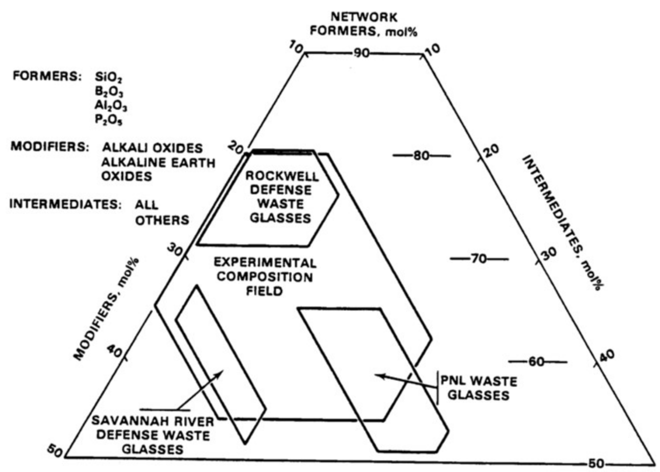


FIGURE 4. Ternary Diagram of Waste Glass Composition Regions

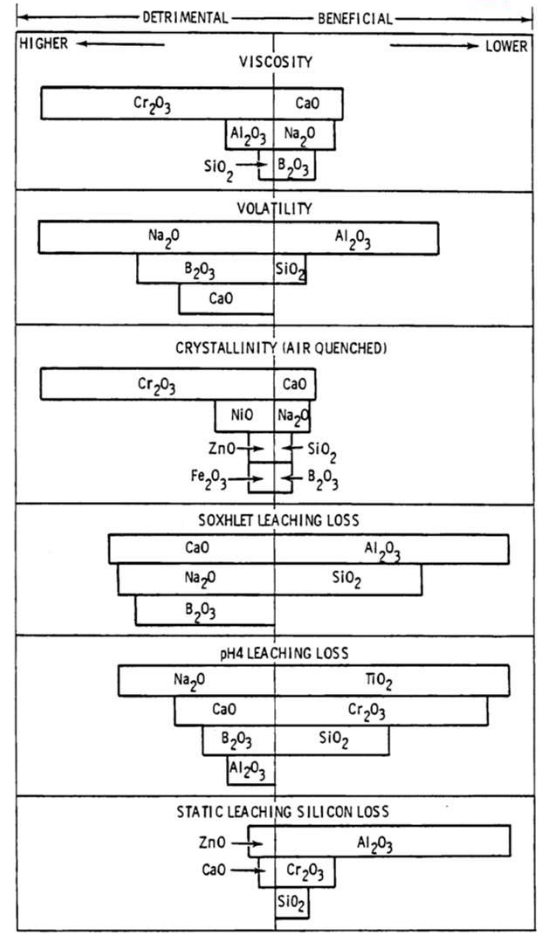
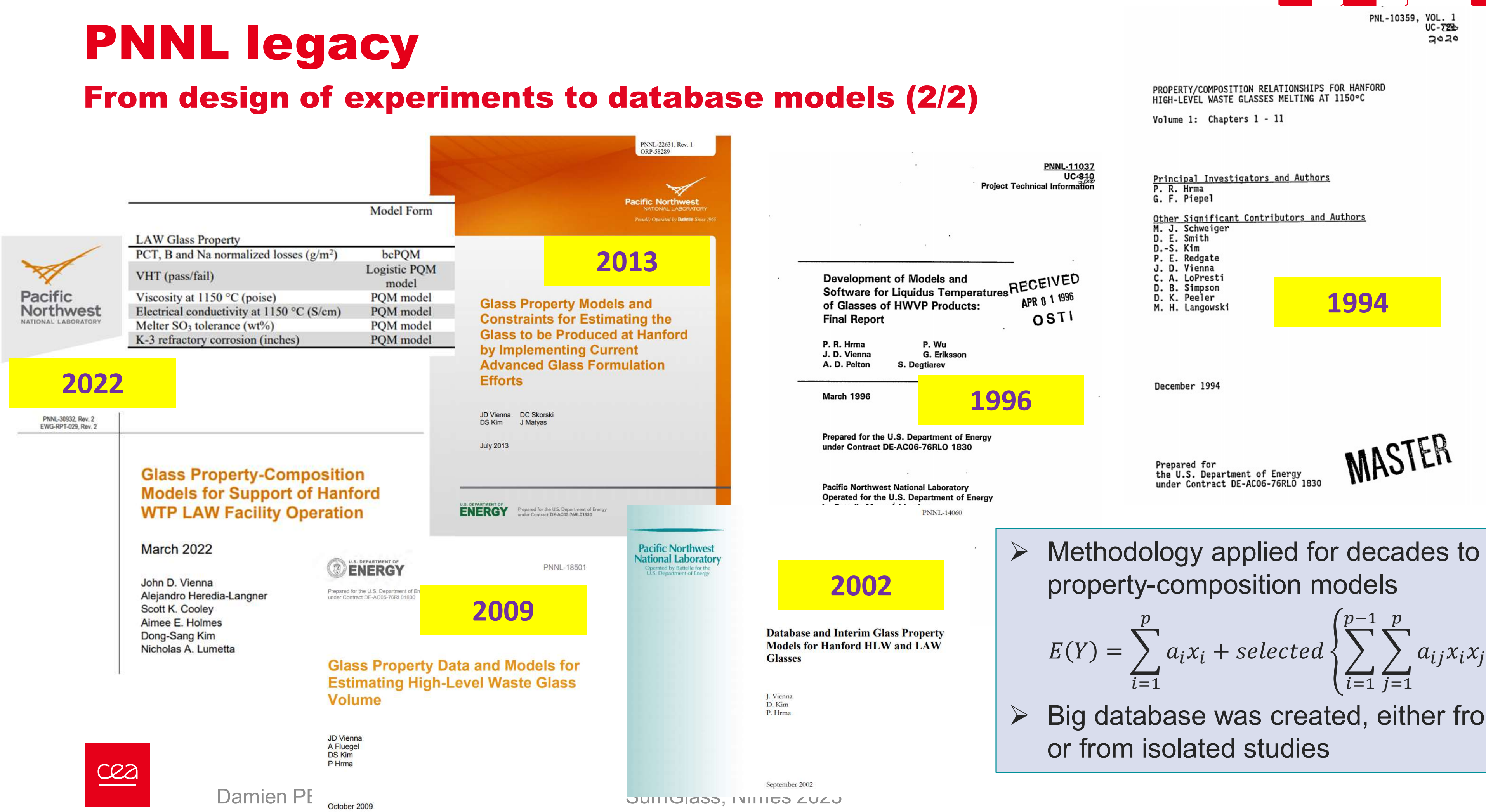


FIGURE 1. Qualitative Summary of Major Component Effects

PNNL legacy

From design of experiments to database models (2/2)



- Methodology applied for decades to build property-composition models

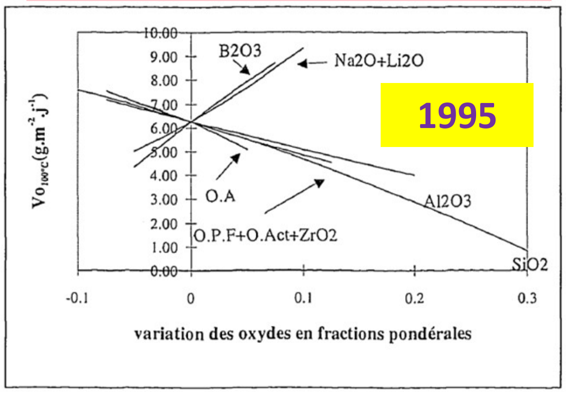
$$E(Y) = \sum_{i=1}^{p} a_i x_i + selected \left\{ \sum_{i=1}^{p-1} \sum_{j=1}^{p} a_{ij} x_i x_j \right\}$$

- Big database was created, either from DOE or from isolated studies

Design of experiments methodology

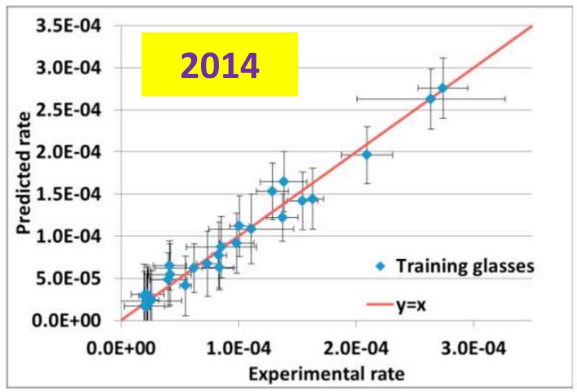
- Some examples of Mixture Designs for waste glass formulation at CEA

Initial dissolution rate (R7T7 glasses)



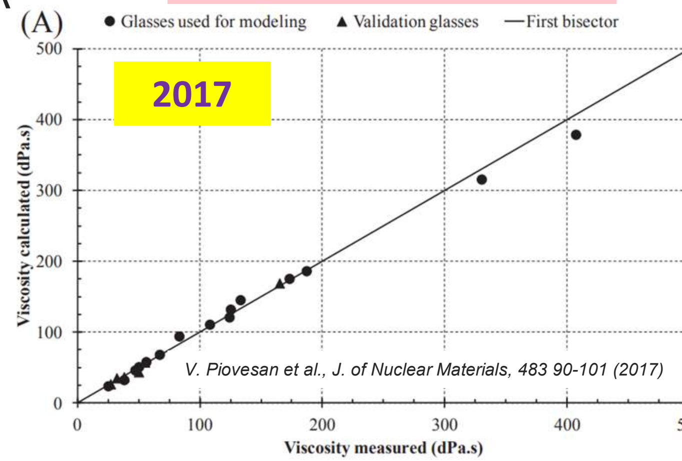
I. Tovena, PhD Thesis, Univ. Montpellier II (1995)
osti.gov/etdeweb/servlets/purl/270252

Residual dissolution rate (R7T7 glasses)

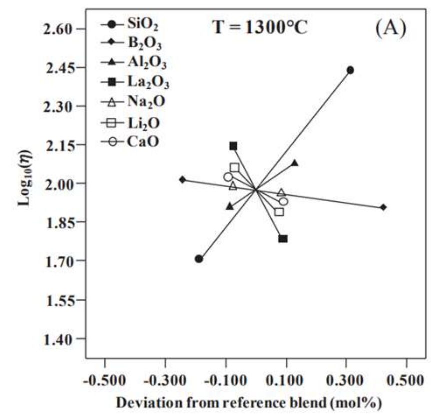


B. Fleury et al., Procedia Materials Science 7 193-201 (2014)

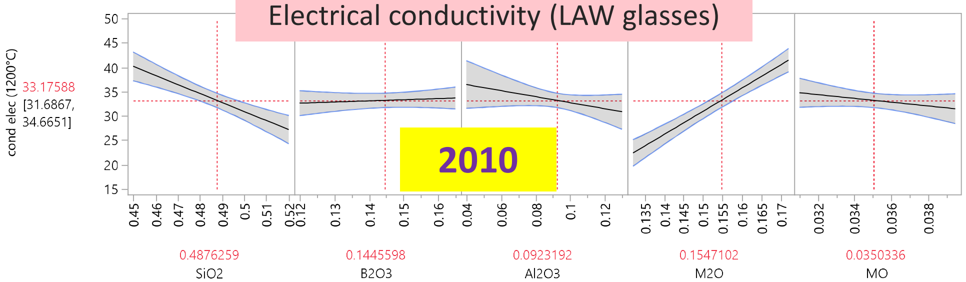
Viscosity (peraluminous glasses)



V. Piovesan et al., J. of Nuclear Materials, 483 90-101 (2017)



Electrical conductivity (LAW glasses)



CEA internal study

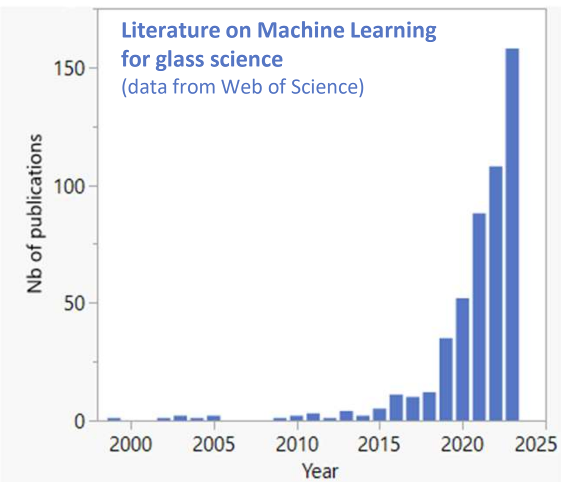
Statistical modeling of glass properties

« Design Of Experiments » (DOE) methodology

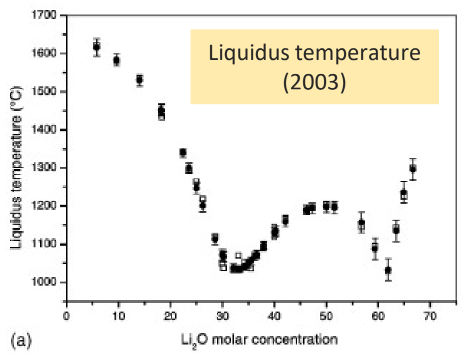
« Machine Learning » (ML) methodology

Machine Learning methodology

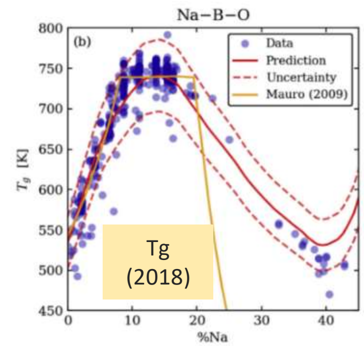
- Some examples of ML use in glass science



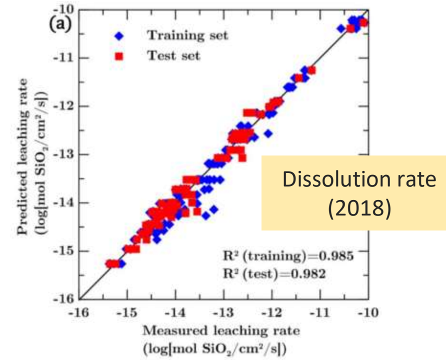
- Development of more and more powerful ML algorithms
- Availability of open access database on glass properties



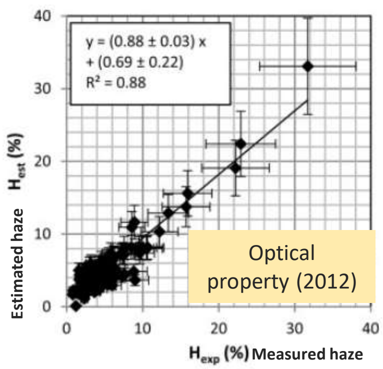
C. Dreyfus, G. Dreyfus / Journal of Non-Crystalline Solids 318 (2003) 63–78



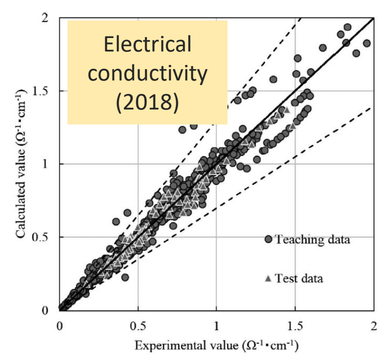
D.R. Cassar et al., Acta Materiala, 159 249-256 (2018)



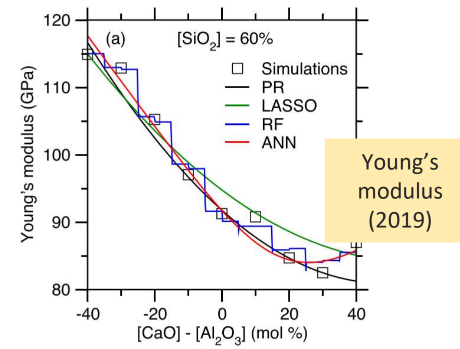
N.M. Anoop Krishnan et al.
Journal of Non-Crystalline Solids 487 (2018) 37–45



A. Verney-Caron et al., Atmospheric Environment, 54 141-148 (2012)



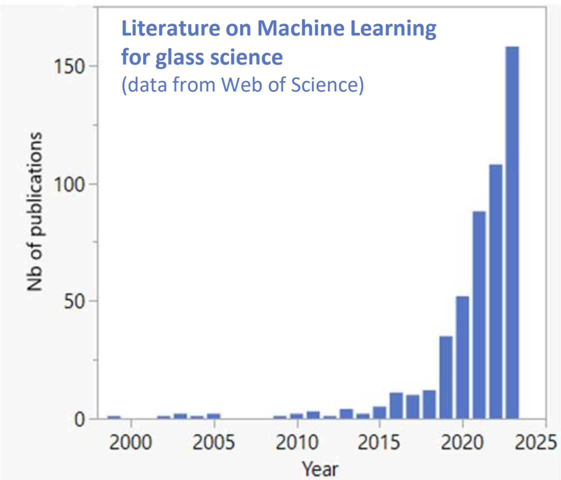
Y. Haraguchi et al., ISIJ International, Vol. 58 (2018), No. 6, pp. 1007–1012



K. Yang et al.., Scientific Reports, 8739, 9 (2019)

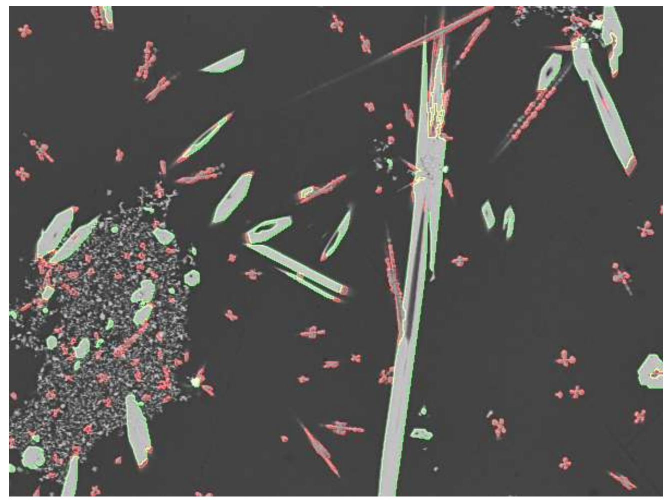
Machine Learning methodology

- Some examples of ML use in glass science



- Development of more and more powerful ML algorithms
- Availability of open access database on glass properties

Example of NN use for SEM image analysis at CEA (crystal recognition)



Test

Superclasse	Measures
Generalized RSquare	0,9647809
Entropy RSquare	0,8870834
RMSE	0,1757289
Mean Abs Dev	0,0775361
Misclassification Rate	0,0348259
-LogLikelihood	199,39034
Sum Freq	1608

Confusion Matrix

Actual	Predicted		
Superclasse	Apatite	Platinoide	Powellite
Apatite	541	1	7
Platinoide	6	527	13
Powellite	1	28	484

Confusion Rates

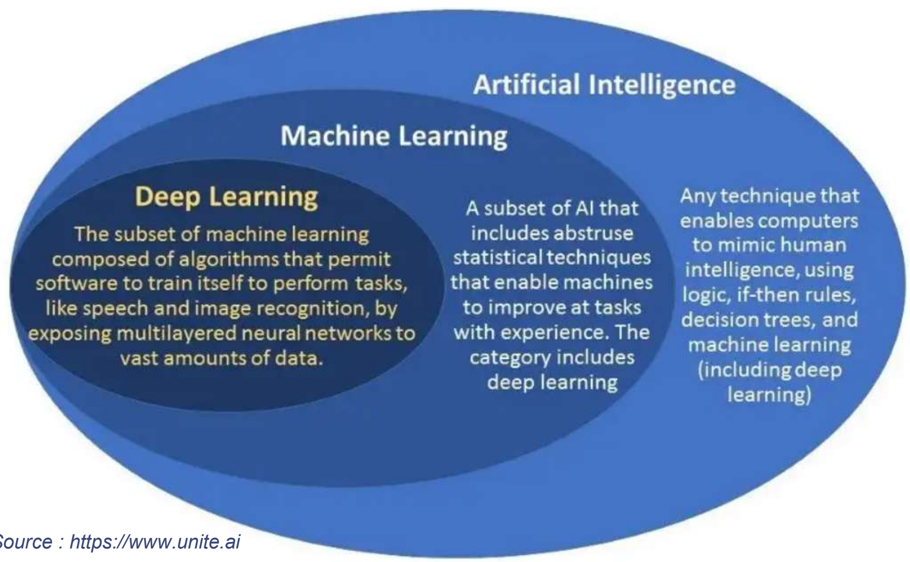
Actual	Predicted		
Superclasse	Apatite	Platinoide	Powellite
Apatite	0,98543	0,00182	0,01275
Platinoide	0,01099	0,96520	0,02381
Powellite	0,00195	0,05458	0,94347

(from internal CEA studies)

- Neural Nets good capability for image analysis
- Rate of classification up to 98%

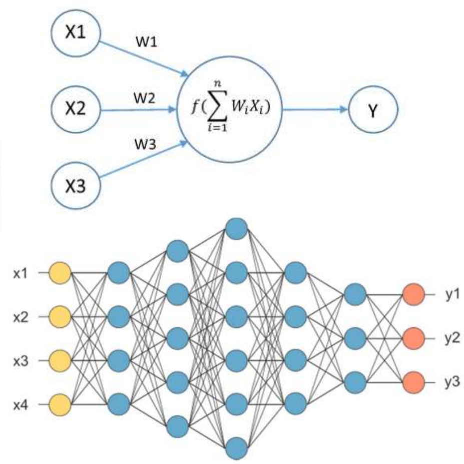
Machine Learning methodology

- Background information

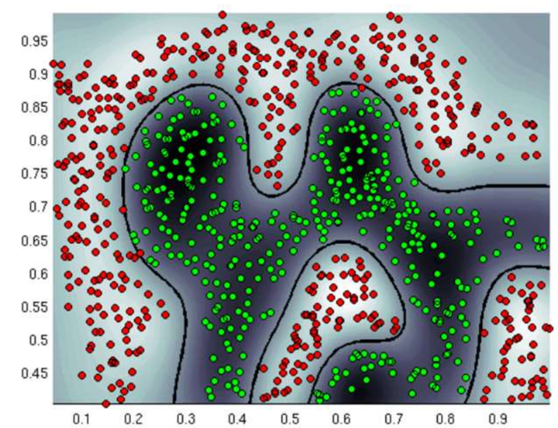


Source : https://www.unite.ai

Neural network schematic description



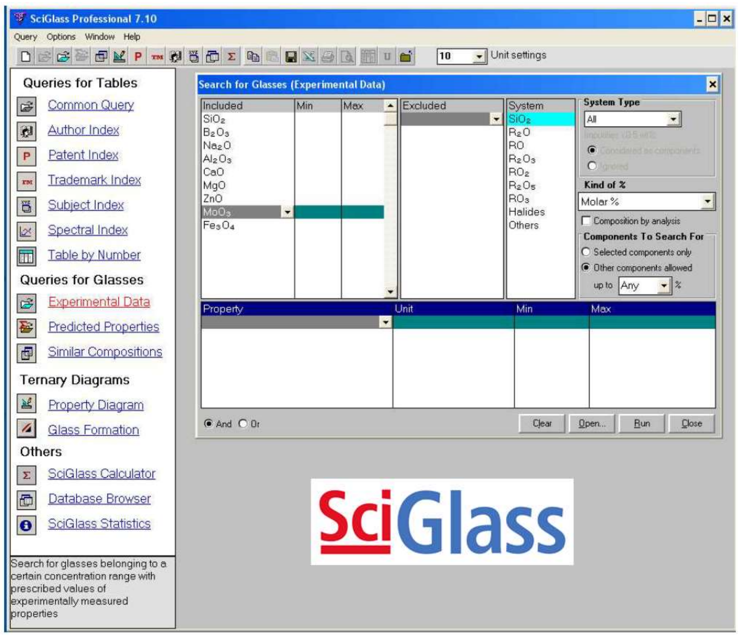
Examples of non-linear methods:
Support vector machines, random forests, neural networks, ...



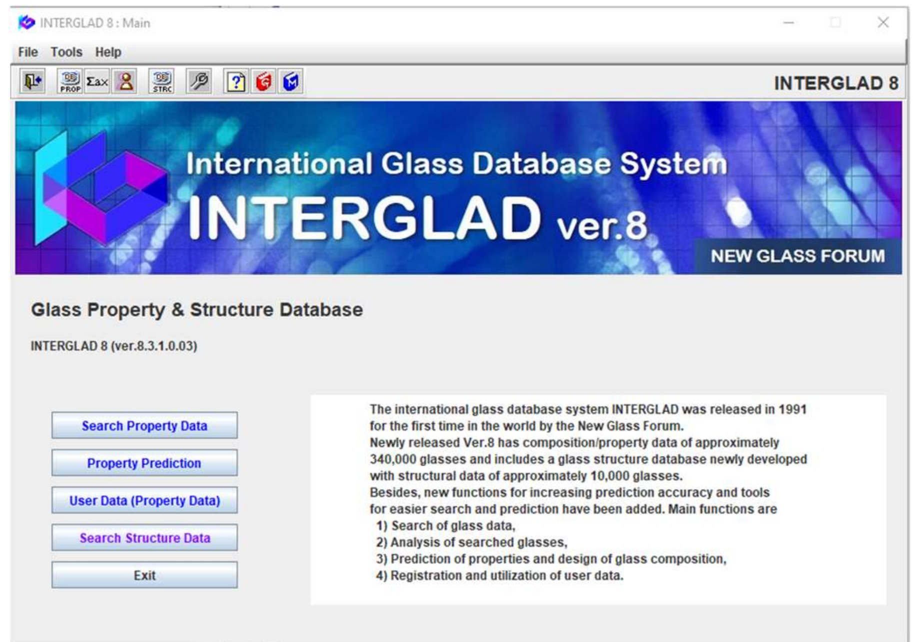
http://openclassroom.stanford.edu/

Machine Learning methodology

- Two main sources of data

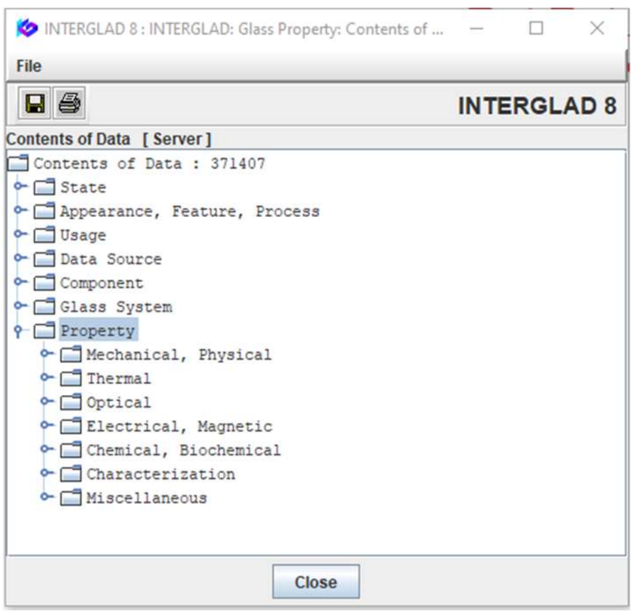


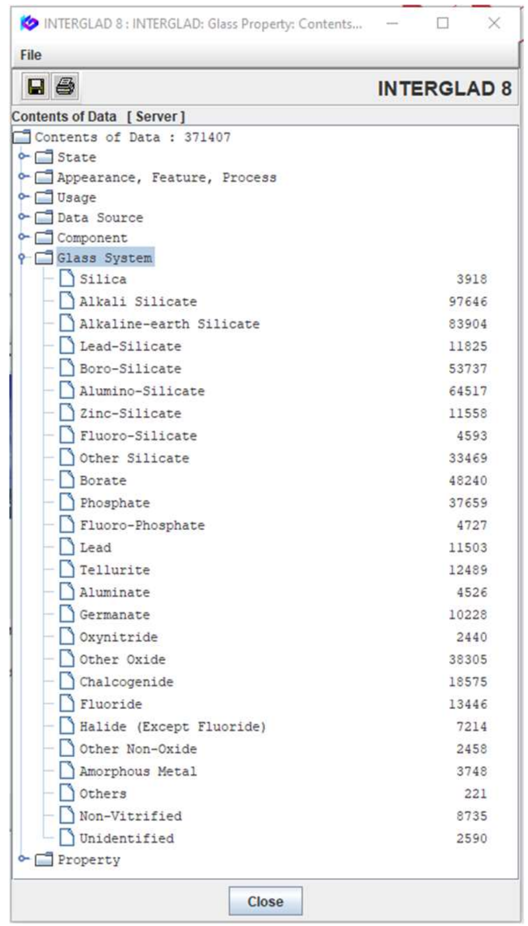
SciGlass database now available under an ODC Open Database License (ODbL) at https://github.com/epam/SciGlass



Machine Learning methodology

- Interglad interface





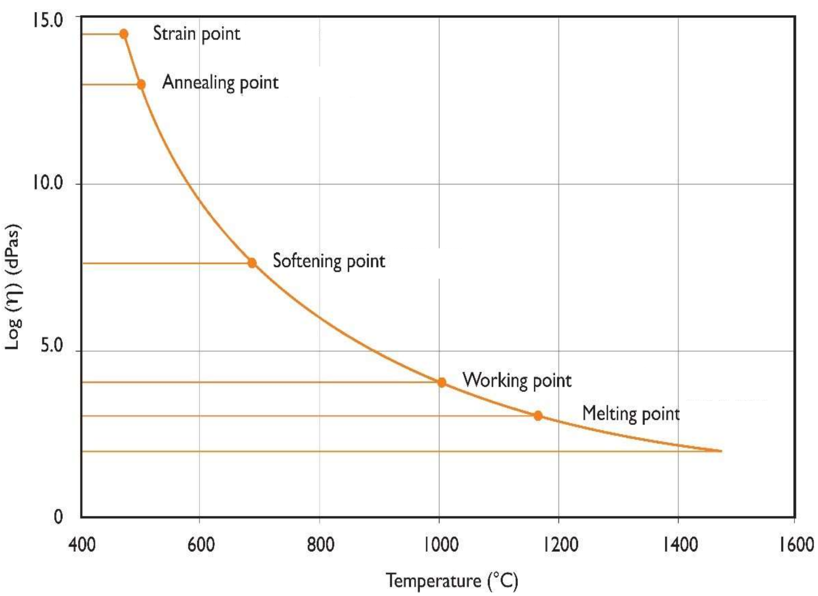
The viscosity case

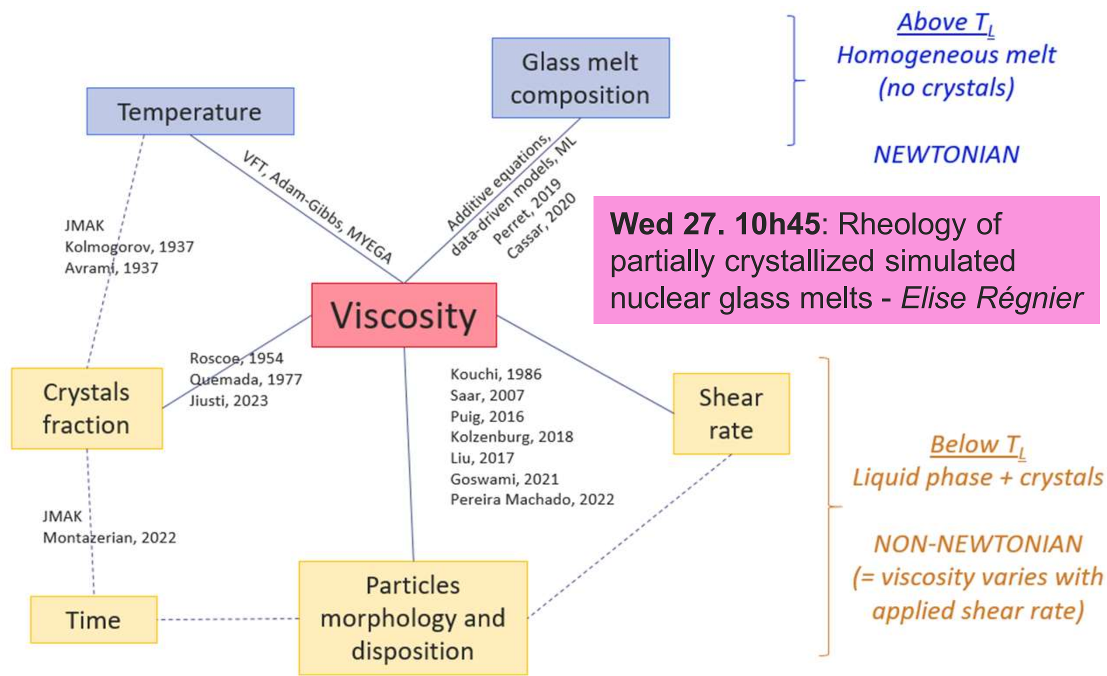
Why is glass melt viscosity prediction so challenging?

→ Range of viscosity values is very wide vs temperature and vs composition (~ 13 orders of magnitude)

→ Viscosity is not simply additive, it depends on bonding forces (valence theory, Myuller) and deformability of unit groups (free volume theory, Turnbull and Cohen) ⇒ chemical dependence of viscosity is extremely complex

→ Viscosity temperature dependence is highly sensitive to phase separation and crystallization



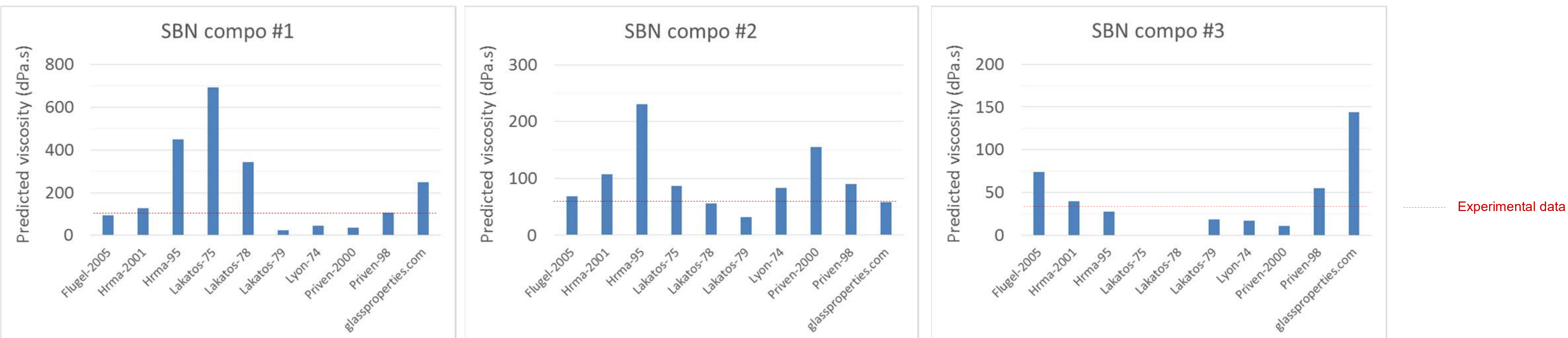


Wed 27. 10h45: Rheology of partially crystallized simulated nuclear glass melts - *Elise Régnier*

The viscosity case

Why is glass melt viscosity prediction so challenging?

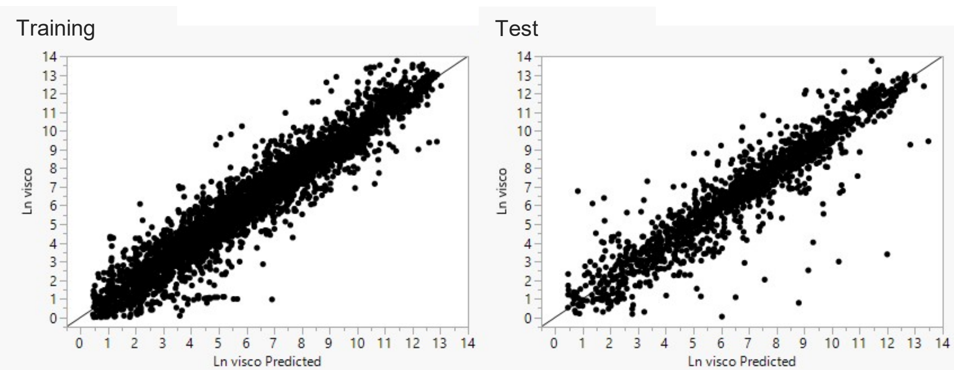
- Viscosity prediction of simple SiO_2-B_2O_3-Na_2O (SBN) glasses at 1200°C



The viscosity case

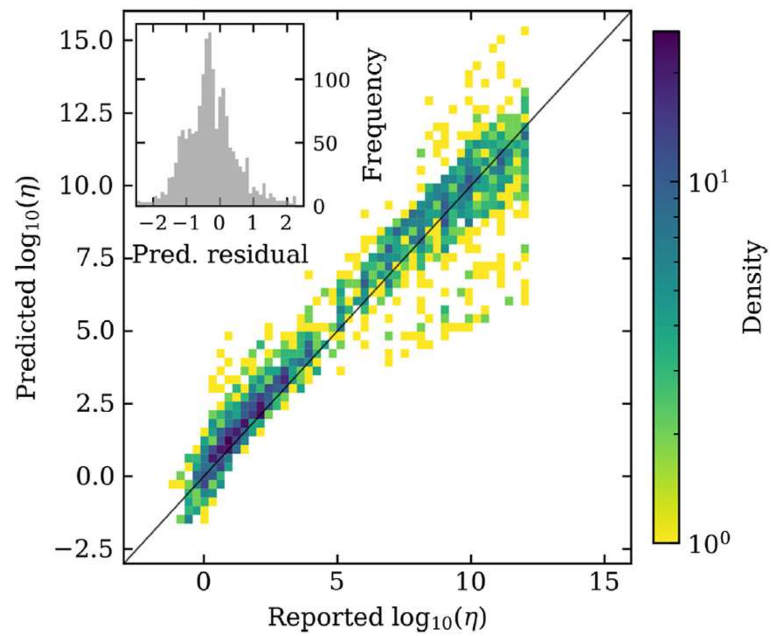
Why is glass melt viscosity prediction so challenging?

- Neural Nets not always appropriate to predict melt viscosity



Ln visco	
Measures	**Value**
RSquare	0.9287269
RMSE	0.7518628
Mean Abs Dev	0.496624
-LogLikelihood	6435.0916
SSE	3208.6293
Sum Freq	5676

Ln visco	
Measures	**Value**
RSquare	0.8822247
RMSE	0.9682362
Mean Abs Dev	0.5912263
-LogLikelihood	2545.9066
SSE	1721.2158
Sum Freq	1836



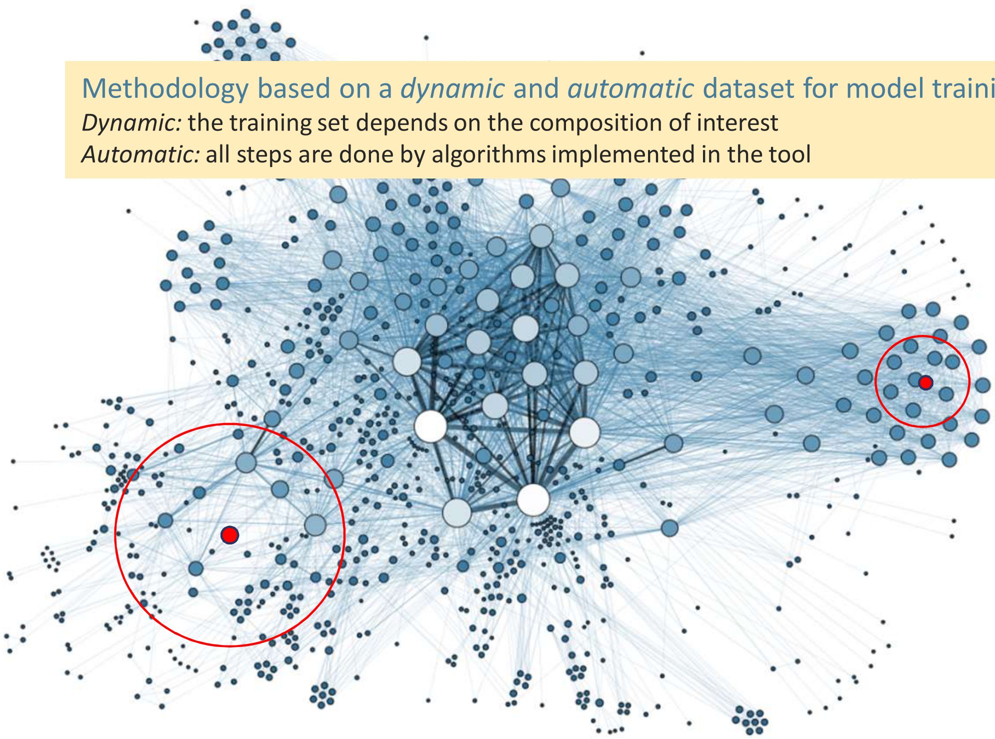
D.R. Cassar, Acta Materialia 206, 116602 (2021)

The viscosity case

Why is glass melt viscosity prediction so challenging?

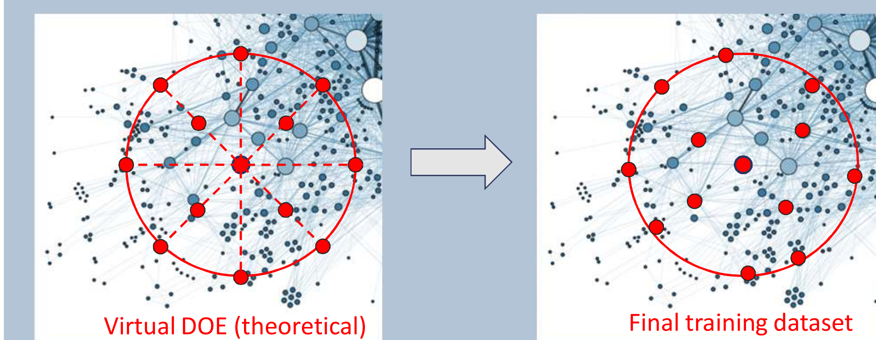
- Innovative approach for glass melt viscosity prediction

(from CEA internal studies)

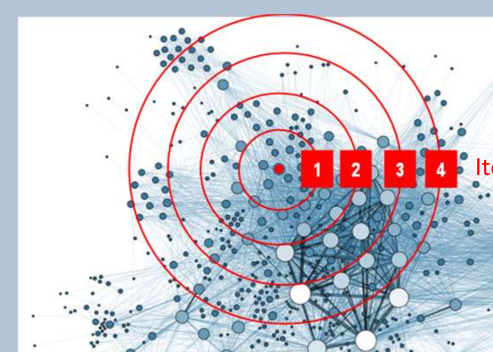


Methodology based on a *dynamic* and *automatic* dataset for model training
Dynamic: the training set depends on the composition of interest
Automatic: all steps are done by algorithms implemented in the tool

Method 1



Method 2



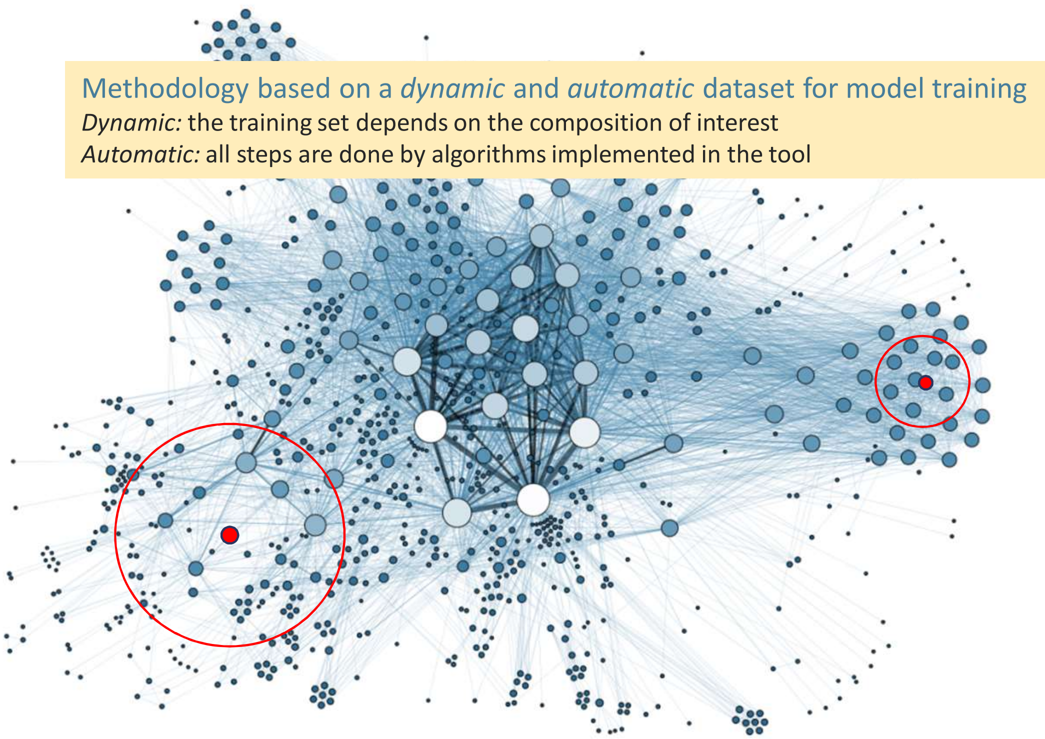
Iterative enlargement

The viscosity case

Why is glass melt viscosity prediction so challenging?

- ❑ Innovative approach for glass melt viscosity prediction

  (from CEA internal studies)



Methodology based on a *dynamic* and *automatic* dataset for model training
Dynamic: the training set depends on the composition of interest
Automatic: all steps are done by algorithms implemented in the tool

For each of the 2 methods:

3 predictive algorithms implemented in the tool
→ Classical polynomial model (MLR)
→ Generalized Regression model
→ Neural Net model

⇒ 6 predicted viscosity values

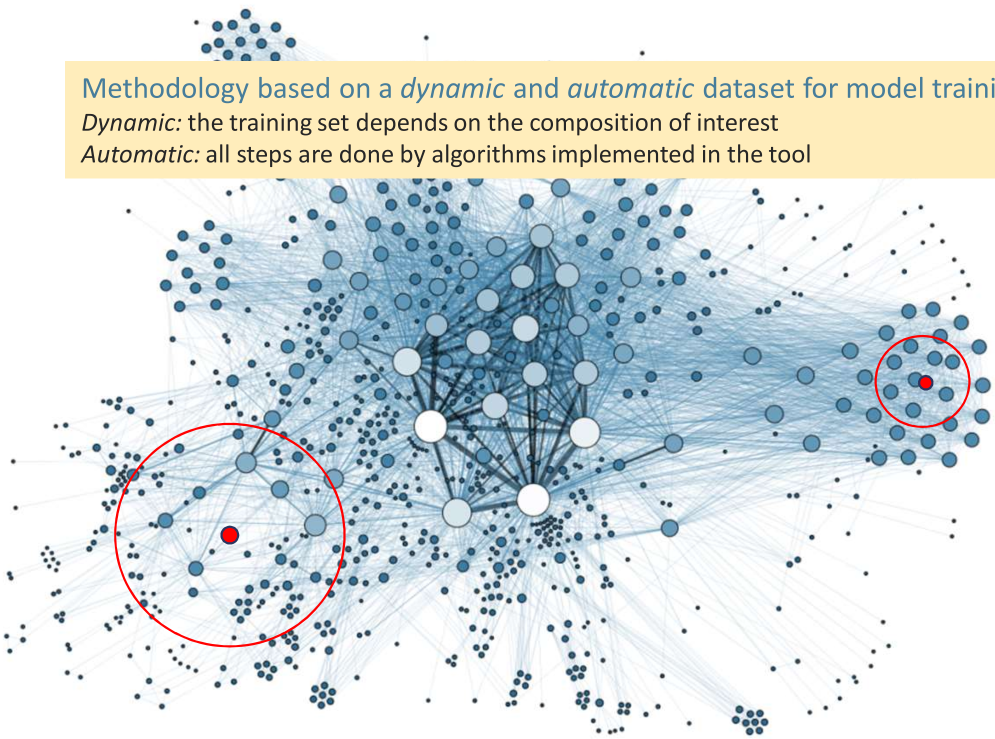
The viscosity case

Why is glass melt viscosity prediction so challenging?

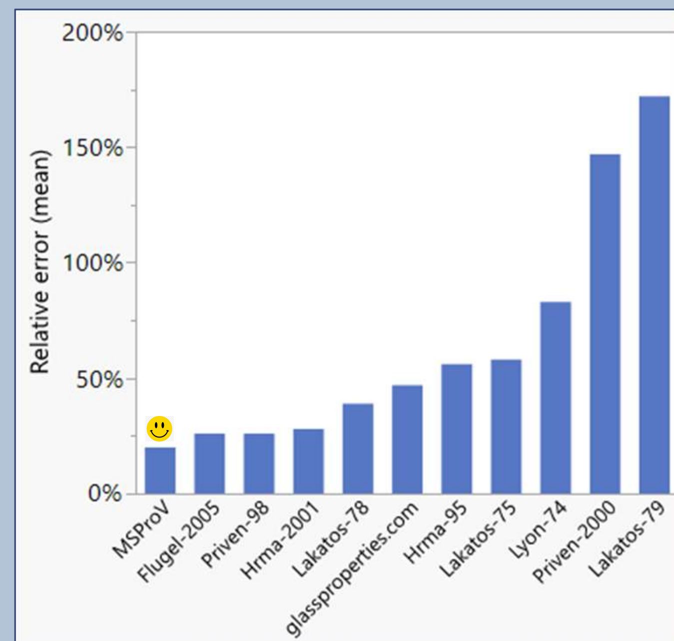
- Innovative approach for glass melt viscosity prediction

(from CEA internal studies)

Methodology based on a *dynamic* and *automatic* dataset for model training
Dynamic: the training set depends on the composition of interest
Automatic: all steps are done by algorithms implemented in the tool



Results obtained on SBN glasses



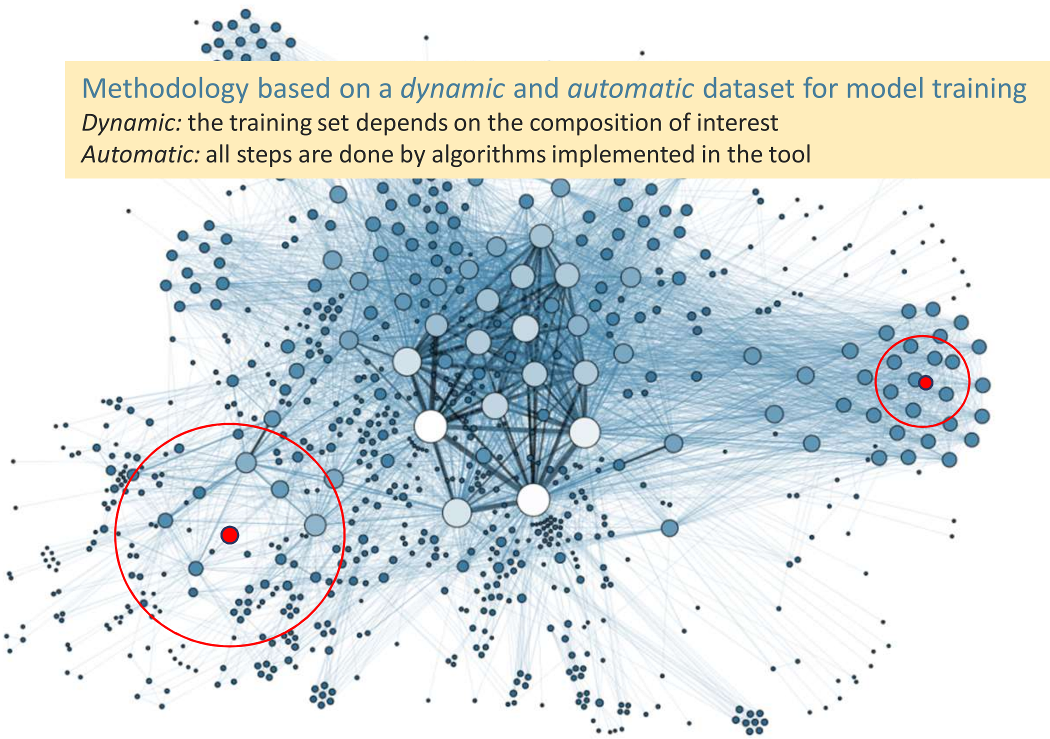
The viscosity case

Why is glass melt viscosity prediction so challenging?

- Innovative approach for glass melt viscosity prediction

(from CEA internal studies)

Methodology based on a *dynamic* and *automatic* dataset for model training
Dynamic: the training set depends on the composition of interest
Automatic: all steps are done by algorithms implemented in the tool



Results obtained on test set (N=230)

Viscosity prediction relative error	Borosilicate glass for nuclear waste	Sodo alumino silica glass	Overall
	N=73	N=55	N=230
Quantile 50% (median)	11%	18%	17%
Quantile 75%	19%	35%	34%
Quantile 90%	37%	73%	77%

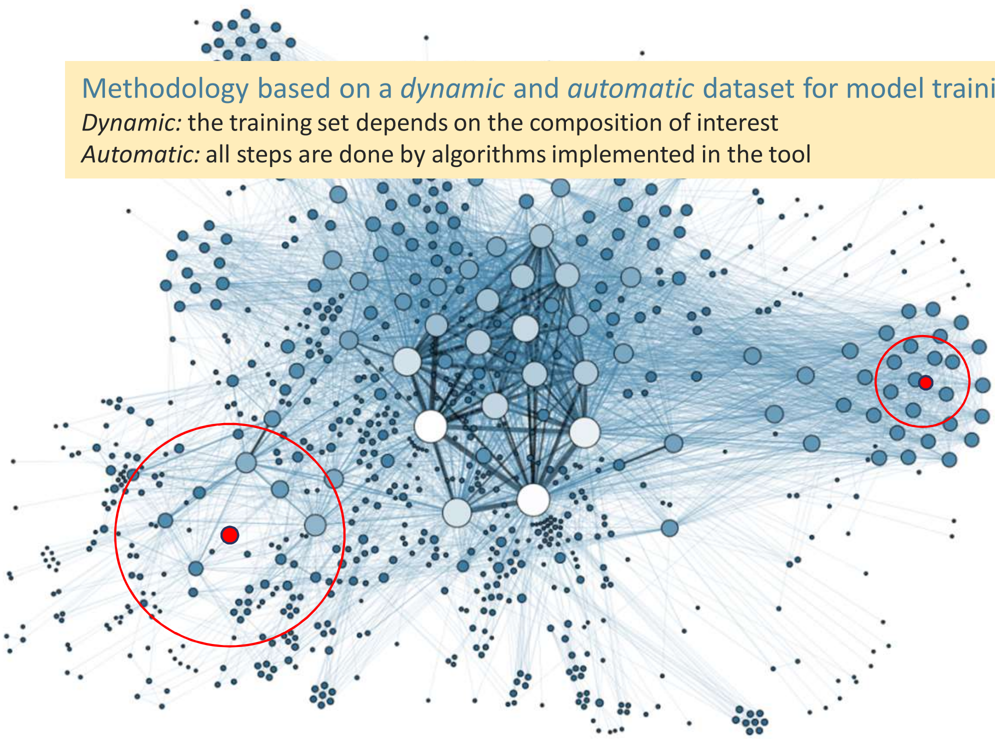
The viscosity case

Why is glass melt viscosity prediction so challenging?

- Innovative approach for glass melt viscosity prediction

  (from CEA internal studies)

Methodology based on a *dynamic* and *automatic* dataset for model training
Dynamic: the training set depends on the composition of interest
Automatic: all steps are done by algorithms implemented in the tool



Results obtained on test set (N=100)
(Tg value prediction)

Tg prediction error	Borosilicate glass		Sodo alumino silica glass		Overall	
	N=80		N=20		N=100	
	Rel.	Abs.	Rel.	Abs.	Rel.	Abs.
Quantile 50% (median)	1,4%	7°C	1,7%	10°C	1,5%	7°C
Quantile 75%	2,8%	13°C	3,2%	17°C	2,9%	14°C
Quantile 90%	5,1%	29°C	4,2%	19°C	4,7%	26°C

Conclusion

- Data-driven models have been used for decades in the field of glass property prediction
- First models were based on the additivity equation and often lead to good prediction accuracy
- Two types of methodology were presented:
  - Experimental designs: very robust and accurate on small domain of composition
    → PNNL legacy from the mid 80s
  - Database and ML: suitable to large datasets
- Glass melt viscosity prediction remains one of the most difficult property to predict on large domain of compositions
- “Black box” Neural Nets not always appropriate to predict melt viscosity from composition only, current limitation of ML not able to take into account mecanisms like crystallization, PGM segregation,…
- Key point in a relevant use of ML relies on the ability to implement glass science expert knowledge in the algorithms

Acknowledgements:

Sophie Schuller, and all the organization and scientific committee of SumGlass 2023

People from DPME/SEME/LFCM, for providing numerous and reliable experimental data for statistical modeling

Thank you for your attention